A Semiparametric Class of Generalized Skew-Elliptical Distributions

Yanyuan Ma* and Marc G. Genton†

January 15, 2003

Abstract

We present a semiparametric representation of generalized skew-elliptical distributions for which the probability density function has the form of a product of an elliptical pdf and a skewing function. By constructing an enumerable dense subset of skewing functions, we are able to define a flexible family of distributions which can capture skewness, heavy tails, and multimodality systematically. It is straightforward to simulate from this family which possesses several invariance properties. We present illustrative examples on fiber-glass and Australian athletes data, as well as on simulated data from a mixture of two normal distributions.

KEY WORDS: Dense subset; Invariance; Kurtosis; Multimodality; Skewness; Stochastic representation.

1 Introduction

Although the normal distribution has played a central role in statistics, there has been recently a growing interest in the construction of flexible parametric non-normal multivariate distributions. The motivation originates from data sets which do not follow the normal law. Such examples typically arise from environmental sciences, finance, engineering, and biomedical ap-

*SAMSI & Center for Research in Scientific Computation, North Carolina State University, Box 8205, Raleigh, NC27695-8205, USA, E-mail: yma@unity.ncsu.edu. The work of Yanyuan Ma is supported by grant NIGMS 1 R01 Gm67299-01.

†Department of Statistics, North Carolina State University, Box 8203, Raleigh, NC 27695-8203, USA. E-mail: genton@stat.ncsu.edu

plications, among others. Departures from the multivariate normal distribution can be achieved by introducing skewness and various levels of kurtosis. One approach in the univariate setting consists of modifying a standard normal random variable through an appropriate nonlinear transformation as suggested by John W. Tukey in 1977, see Hoaglin (1983) for further details and Field and Genton (2002) for an extension to the multivariate case. The resulting g-and-h distribution can be fitted easily with sample quantiles, which is the main advantage of this approach. Probability density functions and moments are more difficult to handle directly.

A second and far more popular approach consists of modifying the probability density function (pdf) of a random vector in a multiplicative fashion. Wang, Boyer, and Genton (2002) showed that any p-dimensional multivariate pdf g admits, for any location parameter $\boldsymbol{\xi} \in \mathbb{R}^p$, a unique skew-symmetric (SS) representation:

$$g(\mathbf{x}) = 2f(\mathbf{x} - \boldsymbol{\xi})\pi(\mathbf{x} - \boldsymbol{\xi}), \tag{1}$$

where $f : \mathbb{R}^p \to \mathbb{R}_+$ is a symmetric pdf and $\pi : \mathbb{R}^p \to [0,1]$ is a skewing function with $\pi(-\mathbf{x}) = 1 - \pi(\mathbf{x})$. By symmetric, we mean $f(\mathbf{x}) = f(-\mathbf{x})$ and we will use "symmetric pdf" and the property $f(\mathbf{x}) = f(-\mathbf{x})$ exchangeably in the sequel. Genton and Loperfido (2002) considered the subfamily of generalized skew-elliptical (GSE) distributions for which the pdf f is elliptically contoured rather than only symmetric. Many definitions of skewed distributions found in the literature are particular generalized skew-elliptical distributions, and thus can be written in the form of a skew-symmetric distribution (1). For instance, Azzalini and Dalla Valle's (1996) multivariate skew-normal distribution corresponds to $f(\mathbf{x}) = \phi_p(\mathbf{x}; \mathbf{0}, \Omega)$ and $\pi(\mathbf{x}) = \Phi(\boldsymbol{\alpha}^T\mathbf{x})$, where $\phi_p(\mathbf{x}; \boldsymbol{\mu}, \Omega)$ is the p-dimensional multivariate normal pdf with mean vector $\boldsymbol{\mu}$ and correlation matrix Ω, Φ is the standard normal cumulative distribution function (cdf), and $\boldsymbol{\alpha}$ is a shape parameter controlling skewness. Similarly, multivariate distributions such as skew-t (Branco and Dey, 2001; Sahu *et al.*, 2001), skew-Cauchy (Arnold and Beaver, 2000), skew-slash (Wang and Genton, 2002), and other skew-elliptical ones (Azzalini and Capitanio, 1999; Branco and Dey, 2001; Sahu *et al.*, 2001) can be represented by the skew-symmetric distribution (1) with appropriate choices of f and π.

In this article, we propose a semiparametric representation of the distributions (1) by constructing an enumerable dense subset of the skewing functions π. The result is a flexible family of distributions which can capture skewness, heavy tails, and multimodality systematically. The construction of the subset is through polynomials, which has a similar flavor as the seminonparametric (SNP) representation proposed by Gallant and Nychka (1987). The latter is defined as the product of the standard normal pdf and the square of a polynomial. One drawback of the SNP pdf is its interpretation. Indeed, the coefficients of the polynomial control departures from normality in a complex fashion and cannot be easily interpreted. For instance, it is not possible to distinguish coefficients that control skewness from coefficients that control tail behavior. Moreover, the coefficients need to be constrained in order to yield a valid SNP density. Another drawback is that a direct procedure for simulation from the SNP pdf is not available, and one needs to resort to rejection sampling. The SNP pdf also has a tendency to produce artifactual waves and modes when the order of the polynomial becomes large. These problems do not occur with our construction.

The content of the paper is organized as follows. In Section 2, we describe a subset of skewing functions based on odd polynomials and prove that it is dense on a compact support. In particular, we define semiparametric skew-normal and skew-t distributions that can have more than one mode, an essential property for some situations. The flexibility and possible multimodality of the new class of distributions is illustrated in Section 3. A stochastic representation useful for simulations and various invariance properties are derived in Section 4. We present three illustrative examples in Section 5, and conclude in Section 6.

2 A dense subset of skewing functions

In this section, we construct a dense subset of skewing functions π. Wang *et al.* (2002, Proposition 2) proved that any skewing function π can be written as:

$$\pi(\mathbf{x}) = H(w(\mathbf{x})), \qquad (2)$$

where $H : \mathbb{R} \to [0, 1]$ is the cdf of a continuous random variable symmetric around 0, and $w : \mathbb{R}^p \to \mathbb{R}$ is an odd function, that is $w(-\mathbf{x}) = -w(\mathbf{x})$. This representation has been used by Azzalini and Capitanio (2002) to define certain distributions by perturbation of symmetry. Note however that the representation (2) is not unique since for *any* cdf H of a continuous random variable symmetric around 0, the corresponding function $w(\mathbf{x}) = H^{-1}(\pi(\mathbf{x}))$ is valid.

Let $P_K(\mathbf{x})$ be an odd polynomial of order K. A polynomial of order K in $\mathbb{R}^p$ is defined as a linear combination of terms of the form $\prod_{i=1}^{p} x_i^{r_i}$, where $k = \sum_{i=1}^{p} r_i \le K$. If each term has an odd order (all k's are odd), then the polynomial is called an odd polynomial, whereas if each term has an even order (all k's are even), it is called an even polynomial. We define semiparametric skew-symmetric (SSS) distributions by modifying (1) into:

$$2f(\mathbf{x}-\boldsymbol{\xi})\pi_K(\mathbf{x}-\boldsymbol{\xi}), \tag{3}$$

where $\pi_K(\mathbf{x}) = H(P_K(\mathbf{x}))$ and H is any cdf of a continuous random variable symmetric around 0. Note that there are no constraints on the coefficients of the polynomial P_K in order to make (3) a valid pdf. In particular, (3) defines semiparametric generalized skew-elliptical (SGSE) distributions when the pdf f is required to be elliptically contoured. For instance, semiparametric generalized skew-normal (SGSN) distributions are defined by:

$$2\phi_p(\mathbf{x};\boldsymbol{\xi},\Omega)\Phi(P_K(A(\mathbf{x}-\boldsymbol{\xi}))), \tag{4}$$

and semiparametric generalized skew-t (SGST) distributions are defined by:

$$2t_p(\mathbf{x};\boldsymbol{\xi},\Omega,\boldsymbol{\nu})T(P_K(A(\mathbf{x}-\boldsymbol{\xi}));\boldsymbol{\nu}), \tag{5}$$

where we use the Choleski decomposition $\Omega^{-1} = A^T A$, t_p denotes a p-dimensional multivariate t pdf, and T denotes a univariate t cdf, both with degrees of freedom $\boldsymbol{\nu}$. Note that we could use Φ, or any other symmetric cdf, instead of T for the skewing function in (5). In practice, a popular choice for the cdf H would be Φ or the univariate cdf corresponding to the symmetric pdf f. Effectively, the following proposition shows that semiparametric skew-symmetric distributions can approximate skew-symmetric distributions arbitrarily well.

Proposition 1. The class of semiparametric skew-symmetric (SSS) distributions is dense in the class of skew-symmetric (SS) distributions on a compact support.

While the result resembles that of the Stone-Weierstrass theory, it is not a direct implication. The proof is given in the Appendix. Proposition 1 shows in particular that the class of generalized skew-elliptical, skew$-t$, and skew-normal distributions can be approximated arbitrarily well by their semiparametric versions.

3 Flexibility and multimodality

In Figure 1, we illustrate the shape flexibility of the SGSN distribution in the univariate case. Its pdf for $K = 3$ is defined by:

$$2\phi_1(x;\xi,\sigma^2)\Phi(\alpha(x-\xi)/\sigma+\beta(x-\xi)^3/\sigma^3). \tag{6}$$

Figure 1(a) depicts the pdf of the SGSN model for $\xi = 0$, $\sigma^2 = 1$, $\alpha = 4$, and $\beta = 0$, i.e. it reduces to Azzalini's (1985) univariate skew-normal distribution. However, when $\beta \neq 0$, the pdf (6) can exhibit bimodality as is shown in Figure 1(b) with $\alpha = 1$, and $\beta = -1$. In general, the higher the degree K of the odd polynomial in the skewing function, the larger is the number of modes allowed in the pdf, thus inducing a greater flexibility in the available shapes. Unfortunately, the number of modes depends on the degree K of the odd polynomial, on the symmetric pdf f, and on the cdf H of the skewing function π_K in a complex fashion. Indeed, even for the univariate situation given by $p = 1$, the modes are determined by zeros of the first derivative of the semiparametric skew-symmetric distribution (3) given by:

$$2f'(x)H(P_K(x)) + 2f(x)H'(P_K(x))P'_K(x), \tag{7}$$

for which the number of zeros cannot be easily computed. Even with restrictions to some specific f and H functions, a general statement on the relation between the number of modes and the order of the polynomial seems not available. However, in the univariate case, if we consider a normal pdf $f = \phi_1$ and a standard normal cdf $H = \Phi$ with an odd polynomial of order $K = 3$, we have the following proposition.

Proposition 2. The class of semiparametric generalized skew-normal (SGSN) distributions with pdf $2\phi_1(x; \xi, \sigma^2)\Phi(\alpha(x-\xi)/\sigma + \beta(x-\xi)^3/\sigma^3)$ has at most 2 modes.

The proof is given in the Appendix. Figure 1 illustrates the result of Proposition 2 by depicting a unimodal and a bimodal pdf from the univariate SGSN with $K = 3$. For $K = 1$, the pdf is always unimodal as was already noted by Azzalini (1985) for the univariate skew-normal distribution.

Next we investigate the flexibility of the SGSN distribution in the bivariate case. Its pdf for $K = 3$, $\boldsymbol{\xi} = \mathbf{0}$, and $\Omega = I_2$ is given by:

$$2\phi_2(x_1, x_2; \mathbf{0}, I_2)\Phi(\alpha_1 x_1 + \alpha_2 x_2 + \beta_1 x_1^3 + \beta_2 x_2^3 + \beta_3 x_1^2 x_2 + \beta_4 x_1 x_2^2). \tag{8}$$

Figure 2 depicts the contours of four different pdfs (8) for various combinations of values of the skewness parameters α_1, α_2, β_1, β_2, β_3, and β_4. In particular, for $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, the pdf is exactly the bivariate skew-normal proposed by Azzalini and Dalla Valle (1996), and known to be unimodal, see Figure 2(a). However, Figures 2(b)-(d) show that many different distributional shapes can be obtained with the parameters $\beta_1, \ldots, \beta_4$, in particular bimodal and

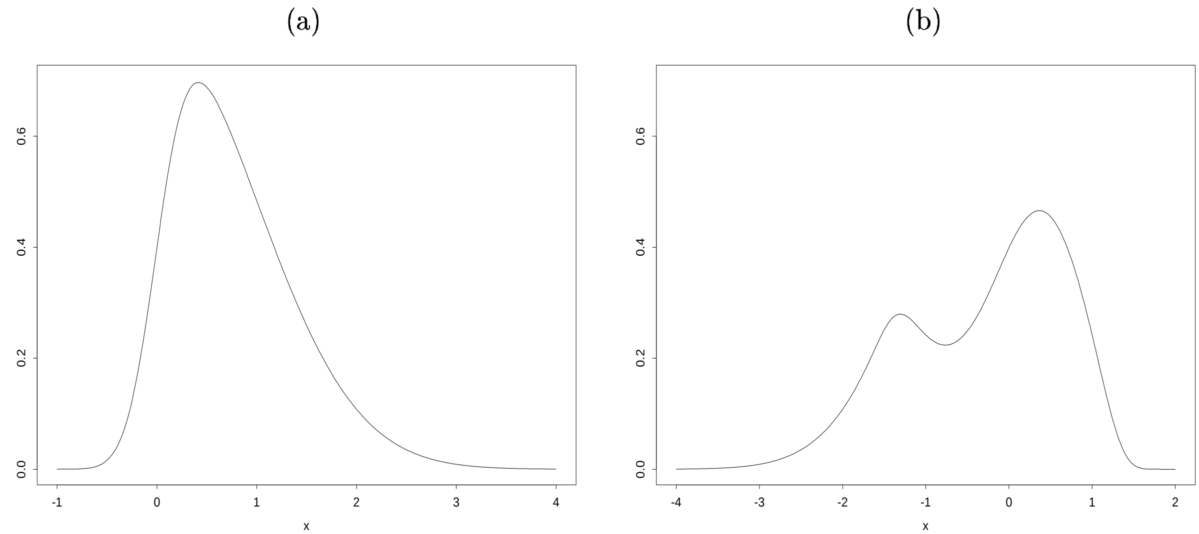

Figure 1: Two members of the univariate SGSN family of distributions with $K = 3$, $\xi = 0$, $\sigma^2 = 1$: (a) $\alpha = 4$, $\beta = 0$ (skew-normal); (b) $\alpha = 1$, $\beta = -1$.

trimodal distributions. Additional flexibility can be imposed on the tail behavior by choosing pdfs other than the normal for the symmetric pdf f, for example a t distribution. This yields semiparametric generalized skew-t distributions (SGST) and will prove useful for applications since they can allow for both fat tails and skewness, see Section 5.

We carry out the estimation and inference for the semiparametric skew-symmetric distribution by maximizing the corresponding likelihood function for a given order K. There are no constraints on the parameters of the skewing function π_K and standard optimization techniques can be used. The order K is chosen adaptively via model selection strategies. Because for a given symmetric pdf f and skewing function π_K the models induced by (3) are nested when K decreases, likelihood ratio tests can be used to identify an appropriate value of K. Model selection criteria such as AIC (twice the loglikelihood minus twice the number of parameters) and BIC (twice the loglikelihood minus the number of parameters times the logarithm of the sample size) can be used as well. In practice, $K = 3$ seems to provide enough flexibility to provide a wide variety of pdfs as is shown in Section 5.

4 Stochastic representation and invariance properties

Modern statistical procedures to derive powerful estimators and tests for complex models, such as the bootstrap and Markov Chain Monte Carlo methods, require the simulation of

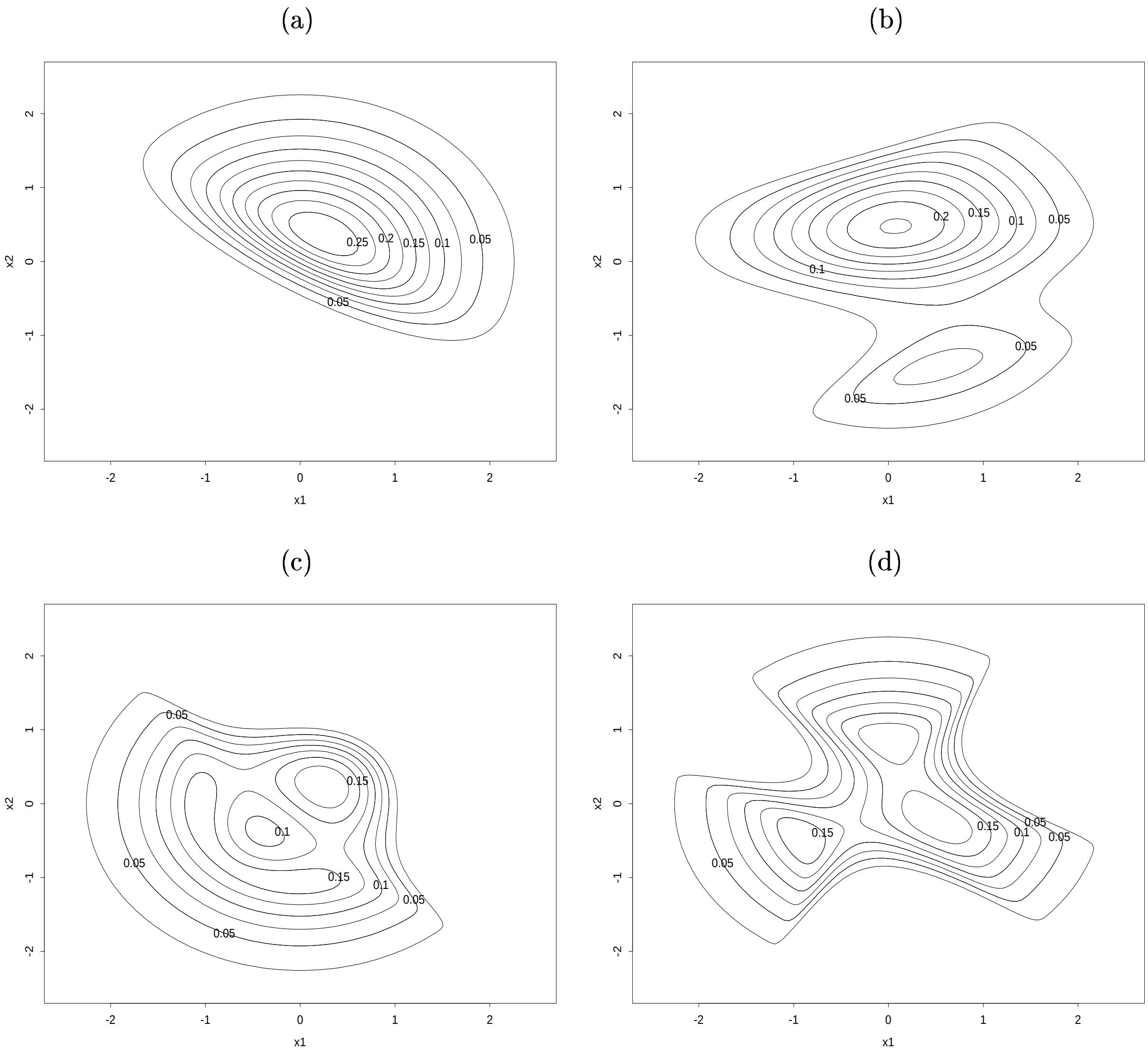

Figure 2: Four members of the bivariate SGSN family of distributions with $K = 3$, $\boldsymbol{\xi} = \mathbf{0}$, and $\Omega = I_2$: (a) $\alpha_1 = 2$, $\alpha_2 = 3$, $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$, $\beta_4 = 0$ (skew-normal; unimodal);
(b) $\alpha_1 = 0$, $\alpha_2 = 2$, $\beta_1 = 0$, $\beta_2 = -1$, $\beta_3 = 0$, $\beta_4 = 1$ (bimodal);
(c) $\alpha_1 = 1$, $\alpha_2 = 1$, $\beta_1 = -2$, $\beta_2 = -2$, $\beta_3 = -1$, $\beta_4 = -1$ (bimodal);
(d) $\alpha_1 = 1$, $\alpha_2 = 0$, $\beta_1 = -1$, $\beta_2 = 2$, $\beta_3 = -4$, $\beta_4 = -1$ (trimodal).

pseudo-realizations from univariate and multivariate distributions. In this section, we first describe a stochastic representation of the semiparametric skew-symmetric distributions (3) which facilitates simple and quick simulation procedures. We then discuss various invariance properties that the family of distributions (3) possesses.

The following proposition describes the stochastic representation of skew-symmetric distributions and therefore also of their semiparametric representations. It provides a constructive description of random vectors having a semiparametric skew-symmetric distribution.

Proposition 3. Let $\mathbf{y}$ be a p-dimensional random vector with symmetric pdf $f(\mathbf{y}-\boldsymbol{\xi})$ and Z be a random variable symmetrically distributed around 0, independent of $\mathbf{y}$, with univariate cdf H. Define $\mathbf{x}$ to be equal to $\mathbf{y}$ conditionally on the event $\{w(\mathbf{y}-\boldsymbol{\xi}) > Z\}$, where w is an odd function. Then $\mathbf{x}$ has the skew-symmetric distribution (1) with the skewing function (2).

The proof of the proposition is given in the Appendix. If w is replaced by an odd polynomial P_K, then $\mathbf{x}$ has the semiparametric skew-symmetric distribution (3). Such constructions were first considered by Azzalini and Dalla Valle (1996) for the multivariate skew-normal distribution with $w(\mathbf{y}) = \boldsymbol{\alpha}^T\mathbf{y}$. Arnold and Beaver (2002) consider also $w(\mathbf{y}) = \alpha_0 + \boldsymbol{\alpha}^T\mathbf{y}$ which is not an odd function and thus cannot be used in our setting. The construction we give above is more general since we consider any odd function w, which can then be approximated arbitrarily closely by an odd polynomial P_K.

In view of the previous discussion, we have the following simple procedure to simulate pseudo-realizations from the semiparametric skew-symmetric distribution (3). First, simulate a random vector $\mathbf{y}$ with pdf $f(\mathbf{y}-\boldsymbol{\xi})$ and a uniform random variable U on the interval $[0,1]$. Then, define $\mathbf{x}$ by:

$$\mathbf{x} = \begin{cases} \mathbf{y} & \text{if } U \le \pi_K(\mathbf{y}-\boldsymbol{\xi}), \\ 2\boldsymbol{\xi}-\mathbf{y} & \text{if } U > \pi_K(\mathbf{y}-\boldsymbol{\xi}). \end{cases} \tag{9}$$

If $f = \phi_p$, $H = \Phi$, $\boldsymbol{\xi} = \mathbf{0}$, and $K = 1$, then the simulation procedure simplifies to:

$$\mathbf{x} = \begin{cases} \mathbf{y} & \text{if } U \le \Phi(\boldsymbol{\alpha}^T\mathbf{y}), \\ -\mathbf{y} & \text{if } U > \Phi(\boldsymbol{\alpha}^T\mathbf{y}). \end{cases} \tag{10}$$

This method thus reduces to the one proposed by Azzalini and Dalla Valle (1996) to simulate from the skew-normal distribution, by noticing that $\Phi^{-1}(U)$ is a standard normal random variable.

The family of semiparametric skew-symmetric distributions possesses interesting invariance properties. For simplicity we assume $\boldsymbol{\xi} = \mathbf{0}$. When $\boldsymbol{\xi} \ne \mathbf{0}$, the following invariance properties hold around $\boldsymbol{\xi}$ instead of $\mathbf{0}$. Wang *et al.* (2002) have proved that if $\mathbf{x}$ is a random vector with a skew-symmetric distribution (1), then the distribution of $\tau(\mathbf{x})$, where τ is an even function, does not depend on the skewing function π. This result extends directly to the case of semiparametric skew-symmetric distributions (3), for any order K of the odd polynomial w. It holds in particular for quadratic forms in the random vector $\mathbf{x}$. As noted by Genton and Loperfido (2002) in the case of generalized skew-elliptical distributions, this property implies that standard inferential methods based on quadratic forms can be misleading. Examples include autocovariance estimation in time series and variogram estimation in spatial statistics, see Genton *et al.* (2001) for further discussions with skew-normal distributions. However, the same property is beneficial for inference from non-random samples, see Genton and Loperfido (2002) for various examples including selection models and case-control data in prospective studies. A particular application of the invariance property implies that even moments are defined by the symmetric pdf f, whereas odd moments depend on the skewing function π_K as well. For instance, if X denotes a random variable having an SGSN distribution (6) with $\xi = 0$ and $\sigma^2 = 1$, then X^2 has a chi-square distribution χ_1^2 for any values of α and β. Visually, this means that the distribution of the square of a random variable with a pdf depicted in Figure 1(a) or 1(b) is the same χ_1^2. Similarly, the distributions of $\mathbf{x}^T\mathbf{x}$ is χ_2^2 for $\mathbf{x}$ having any of the four pdfs described in Figure 2. The odd moments of semiparametric skew-symmetric distributions depend however on the skewing function and analytical expressions seem not to be available except for very simple cases such as Azzalini and Dalla Valle's (1996) skew-normal distribution.

5 Examples

In this section, we present three applications of semiparametric generalized skew-elliptical distributions. The first one is concerned with a unidimensional data set of breaking strengths values of 1.5cm long glass fibers. Jones and Faddy (2002) and Azzalini and Capitanio (2002) fit two forms of skew-t distributions to these data. They both noted skewness on the left as well as heavy tail behavior. We fit a semiparametric generalized skew-t (SGST) distribution (5) with pdf:

$$2t_1(x;\xi,\sigma^2,\nu)T(P_K((x-\xi)/\sigma);\nu), \tag{11}$$

for $K = 1$ and $K = 3$. The fitted parameters, obtained by maximizing the corresponding semiparametric likelihood function, are listed in Table 1. Note the small values for $\hat{\nu}$ indicating tails much heavier than the normal distribution. The fitted pdfs are depicted in Figure 3 for $K = 1$ (solid line) and $K = 3$ (dotted line), along with a histogram of the fiber-glass data. There appear to be not too much difference between the two models. Because the model for

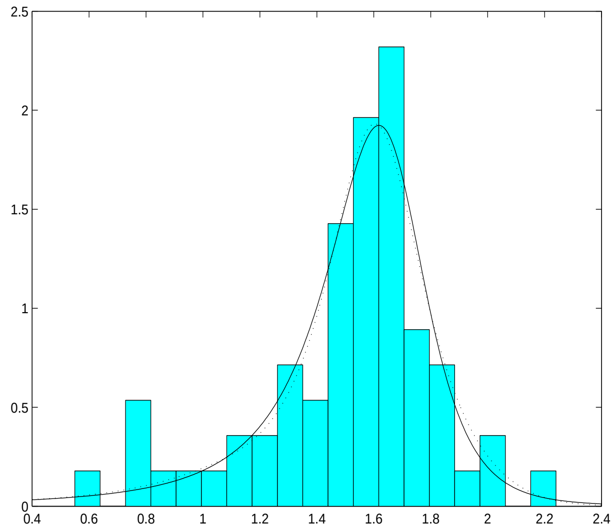

Figure 3: Histogram of the fiber-glass data and fitted pdfs of the SGST model with $K = 1$ (solid line) and $K = 3$ (dotted line) by semiparametric maximum likelihood.

$K = 1$ is nested in the one for $K = 3$, we can use a likelihood ratio test (LRT) for the null hypothesis $H_0 : \beta = 0$ with the approximate asymptotic distribution χ_1^2 or use the AIC or BIC criteria. The results are tabulated in Table 2. All three methods favor the SGST model with $K = 1$ which does not allow for bimodality.

The second example is the Australian athletes data set discussed by Azzalini and Dalla Valle (1996). It consists of several variables measured on 202 athletes and we focus on height (Ht) and body mass index (BMI). We fit a bivariate semiparametric generalized skew-normal (SGSN) distribution (4) with $K = 1$ and $K = 3$. The parameters, estimated by maximizing the corresponding semiparametric likelihood function, are listed in Table 3. The contours of the fitted bivariate pdfs are depicted in Figure 4(a) for $K = 1$ and in Figure 4(b) for $K = 3$. A likelihood ratio test for the null hypothesis $H_0 : \beta_1 = \ldots = \beta_4 = 0$, using the approximate

Table 1: Fitted values of the univariate SGST model for $K = 1$ and $K = 3$ on the fiber-glass data.

	$\hat{\xi}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\nu}$
$K = 1$	1.67	5.14	-0.60	—	2.05
$K = 3$	1.60	5.43	0.07	-0.04	2.11

Table 2: Model selection criteria for $K = 1$ and $K = 3$ on the fiber-glass data.

	LRT (p-value)	AIC	BIC
$K = 1$	—	-31.9	-40.4
$K = 3$	0.42	-33.2	-43.9

Table 3: Fitted values of the bivariate SGSN model for $K = 1$ and $K = 3$ on the Australian athletes Ht/BMI data.

	$\hat{\xi}_1$	$\hat{\xi}_2$	$\hat{a}_{11}$	$\hat{a}_{12}$	$\hat{a}_{22}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$K = 1$	180.51	19.98	0.011	-0.005	0.06	-0.83	3.11	—	—	—	—
$K = 3$	180.87	24.40	0.012	-0.012	0.11	0.21	-1.26	0.18	0.16	-0.18	-0.76

asymptotic distribution χ^2_4 is used as well as the AIC and BIC criteria. The results are tabulated in Table 4. All three methods suggest that $K = 3$ is a better model than $K = 1$. We further test for $K = 5$. The result of the likelihood ratio test relative to $K = 3$, as well as the AIC and BIC criteria, are shown in Table 4. Although the p-value of the LRT and the AIC score

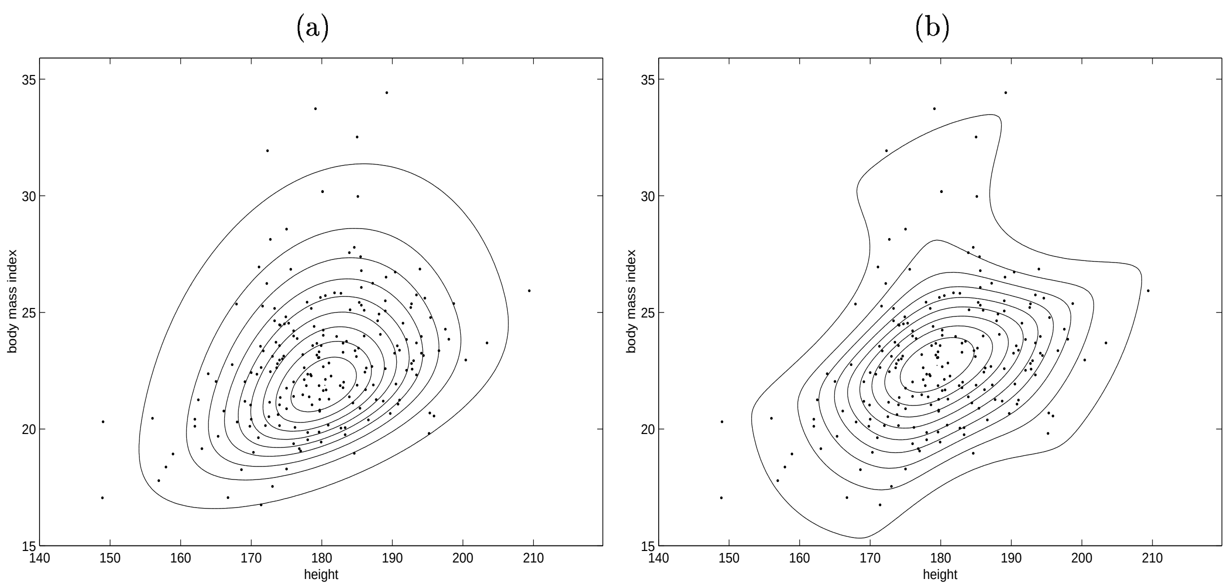

Figure 4: Bivariate contours of the fitted SGSN model to the Australian athletes Ht/BMI data by semiparametric likelihood: (a) $K = 1$; (b) $K = 3$.

indicate that $K = 5$ is a better fit for the data, BIC suggests that $K = 5$ imposes too much model complexity for the gain. We decide to adopt a more complex model only when all three methods indicate so, hence we keep $K = 3$ as our final model. Notice that the final model is unimodal and reveals a complex shape that cannot be captured by a skew-normal distribution.

Table 4: Model selection criteria for $K = 1, 3, 5$ on the Ht/BMI data.

	LRT (p-value)	AIC	BIC
$K = 1$	—	-1103.6	-1126.7
$K = 3$	0.0002	-1089.0	-1125.4
$K = 5$	0.0119	-1084.7	-1140.9

The third example is a simulated data set of size 100 from a mixture of two bivariate normal distributions:

$$(1 - \varepsilon)N_2(\boldsymbol{\mu}_1, \Sigma_1) + \varepsilon N_2(\boldsymbol{\mu}_2, \Sigma_2), \tag{12}$$

with $\varepsilon = 0.4$, $\boldsymbol{\mu}_1 = (0, 0)^T$, $\boldsymbol{\mu}_2 = (5, 4)^T$, and:

$$\Sigma_1 = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}. \tag{13}$$

Figure 5(a) depicts the 100 simulated data along with the bivariate contours of the pdf corresponding to (12), which shows bimodality. We fit a bivariate semiparametric generalized skew-normal (SGSN) distribution (4) with $K = 1$ and $K = 3$. The parameters, estimated by maximizing the corresponding semiparametric likelihood function, are listed in Table 5. The contours of the fitted bivariate pdfs are depicted in Figure 5(b) for $K = 1$ and in Figure 5(c) for $K = 3$. The case $K = 1$ corresponds to Azzalini and Dalla Valle's (1996) bivariate skew-normal

Table 5: Fitted values of the bivariate SGSN model for $K = 1$ and $K = 3$ on the simulated mixture data.

	$\hat{\xi}_1$	$\hat{\xi}_2$	$\hat{a}_{11}$	$\hat{a}_{12}$	$\hat{a}_{22}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$K = 1$	-1.64	-1.93	0.36	-0.29	0.28	2.42	6.84	—	—	—	—
$K = 3$	1.72	2.08	0.42	-0.26	0.27	-0.78	-1.60	0.21	0.40	0.31	0.56

distribution, which of course cannot capture the bimodality. The fit with $K = 3$ captures the bimodality and adapts closely to the shape of the simulated data. We test the model with the likelihood ratio test and AIC, BIC criteria, and find that all three select the model with $K = 3$. We further fit a model with $K = 5$ and find that the likelihood ratio test and AIC criterion favor $K = 5$ to $K = 3$, while BIC favors the model with $K = 3$. Again, the decision whether to accept $K = 3$ or $K = 5$ depends on how much one favors each method. For the same reason as in the previous example, we keep the SGSN model with $K = 3$. The results of the model selection are tabulated in Table 6.

In Figure 6 we plot the marginal pdfs of the mixture model (12) (solid line) and of the fitted SGSN (dotted line) for $K = 1$ in panels (a) and (b), and for $K = 3$ in panels (c) and (d). Note in panel (a) that the fitted marginal (dotted line) cannot capture the bimodality, whereas it is captured by the model with $K = 3$ in panel (c).

Table 6: Model selection criteria for $K = 1, 3, 5$ on the simulated mixture data.

	p-value	AIC	BIC
$K = 1$	—	-898.7	-916.9
$K = 3$	0.0009	-888.0	-916.7
$K = 5$	0.0085	-882.8	-927.0

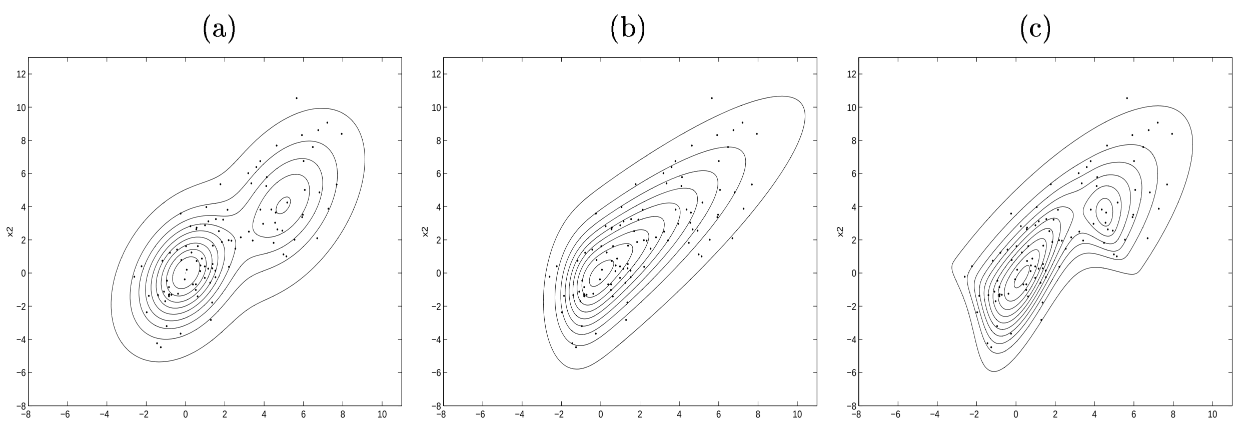

Figure 5: Simulated data set of size 100 from a mixture of two normal distributions, with contours of the corresponding bivariate pdf in panel (a). Contours of the bivariate fitted pdf from the SGSN model with $K = 1$ in panel (b) and with $K = 3$ in panel (c).

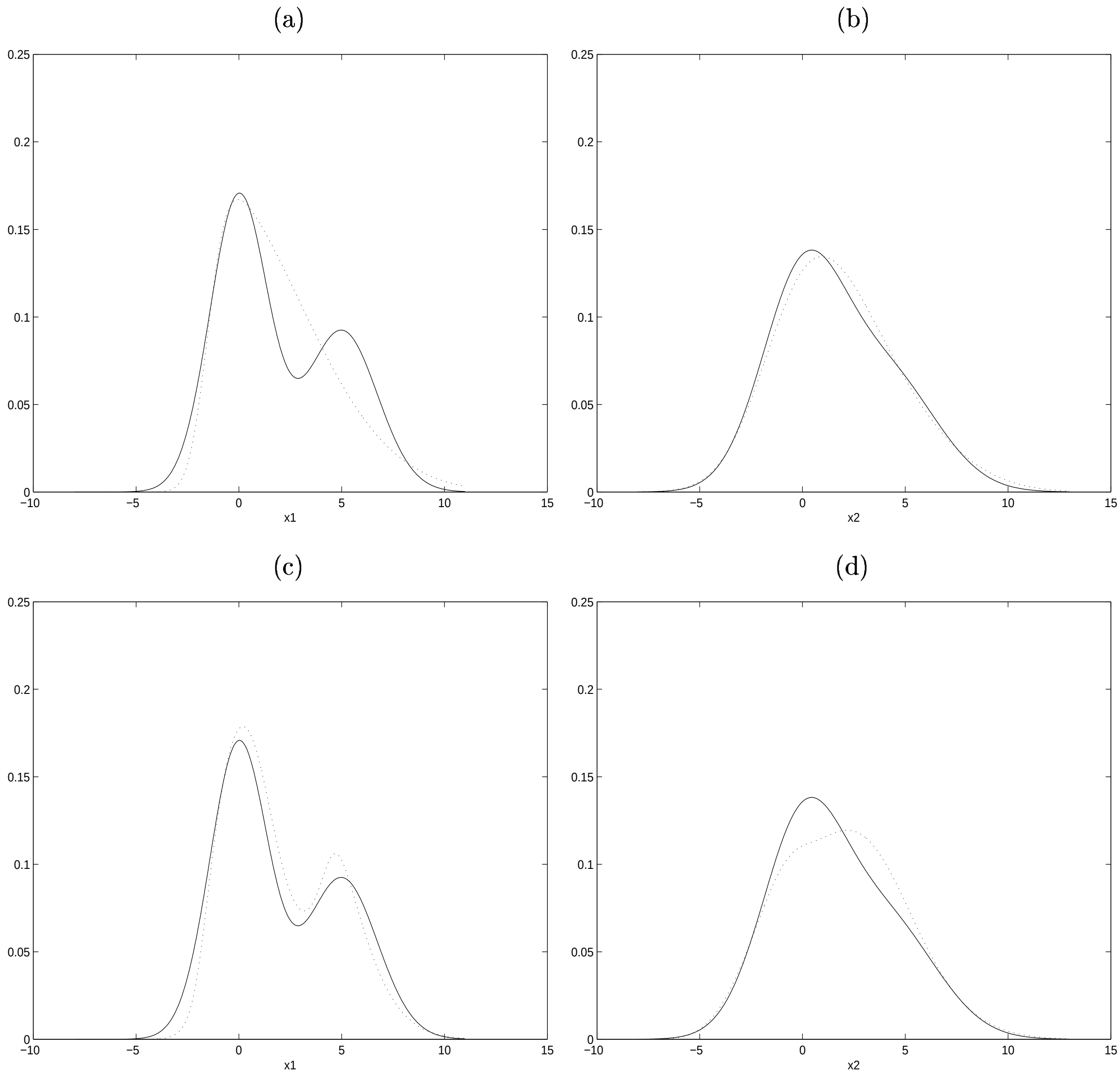

Figure 6: Marginal pdfs for a mixture of two normal distributions (solid line) and for fitted SGSN model (dotted line). Panel (a) and (b): $K = 1$. Panel (c) and (d): $K = 3$.

6 Discussion

We have introduced semiparametric skew-symmetric distributions, a flexible class that can take skewness, heavy tails, and multimodality, into account. It is based on an approximation of the skewing function by a univariate symmetric cdf evaluated at odd polynomials. Using the Stone-Weierstrass theorem, we have proved that this approximation can be made arbitrarily accurate by increasing the order of the odd polynomial. However, the number of coefficients of the polynomial increases quickly with its order K, especially so when the dimension p of the distribution is large as well. In this case, Monte Carlo Markov Chain methods might be more appropriate than direct likelihood maximization. In light of the examples in Section 5, it seems that $K = 3$ is sufficient for practical applications.

The choice of the symmetric pdf f is of practical importance. Although it has been shown that any multivariate pdf g can be represented by a skew-symmetric distribution, a parametric form for the pdf f needs to be specified for our applications. It turns out that the normal and the t pdfs are the most natural ones, yielding the semiparametric models SGSN and SGST. The later model is particularly well suited to capture heavy tails, possibly Cauchy-like, in addition to skewness and multimodality. More sophisticated symmetric pdfs f could be used as well, e.g. such as the slash distribution. The choice of the cdf H has theoretically no impact, but in practice it can influence the order K of the polynomial P_K. For applications, H should be chosen to facilitate computations. A natural choice is $H = \Phi$ or the cdf corresponding to the symmetric pdf f. Finally, it is worth mentioning that our methodology can also be used for proportion data with skewness, see Wang *et al.* (2002) for an illustrative example with $K = 1$.

7 Appendix

7.1 Proof of Proposition 1

We first state the classical Stone-Weierstrass Theorem here for completeness, and refer to Rudin (1973) for a detailed account.

Theorem 1 (Stone-Weierstrass) *If X is a compact Hausdorff space with at least two points and A is a subalgebra of the Banach algebra $C(X)$ which separates points and contains a non-zero constant function, then A is dense in $C(X)$.*

A direct result from Theorem 1 is that the family of all the polynomials on a compact subset of $\mathbb{R}^p$ form a dense subset of all the continuous functions on the same compact subset. This is a generalization of the celebrated Weierstrass theorem in the one dimensional case.

Corollary 1 (Weierstrass in $\mathbb{R}^p$) *Suppose h is a continuous real function defined on a compact subset S of $\mathbb{R}^p$. For an arbitrary $\epsilon > 0$, there exists a real polynomial P in $\mathbb{R}^p$ such that for all $\mathbf{x} \in S$, we have $|h(\mathbf{x}) - P(\mathbf{x})| < \epsilon$.*

We use Corollary 1 to prove the following statements:

(a) Suppose w is a continuous odd real function defined on a compact subset S of $\mathbb{R}^p$. For an arbitrary $\epsilon > 0$, there exists an odd real polynomial P_o in $\mathbb{R}^p$ such that for all $\mathbf{x} \in S$, we have $|w(\mathbf{x}) - P_o(\mathbf{x})| < \epsilon$.

(b) Similarly, suppose v is a continuous even real function defined on a compact subset S of $\mathbb{R}^p$. For an arbitrary $\epsilon > 0$, there exists an even real polynomial P_e in $\mathbb{R}^p$ such that for all $\mathbf{x} \in S$, we have $|v(\mathbf{x}) - P_e(\mathbf{x})| < \epsilon$.

To prove (a), first note that from Corollary 1, there exists a polynomial P such that $|w(\mathbf{x}) - P(\mathbf{x})| < \epsilon$ for any $\mathbf{x} \in S$. We decompose P into an even term P_e and an odd term P_o, i.e. $P = P_e + P_o$. Then $|w(\mathbf{x}) - P_e(\mathbf{x}) - P_o(\mathbf{x})| < \epsilon$ and $|w(-\mathbf{x}) - P_e(-\mathbf{x}) - P_o(-\mathbf{x})| < \epsilon$. Because w and P_o are odd, and P_e is even, we get $| - w(\mathbf{x}) - P_e(\mathbf{x}) + P_o(\mathbf{x})| < \epsilon$. Notice that $2|w(\mathbf{x}) - P_o(\mathbf{x})| \leq |w(\mathbf{x}) - P_e(\mathbf{x}) - P_o(\mathbf{x})| + |-w(\mathbf{x}) - P_e(\mathbf{x}) + P_o(\mathbf{x})| < 2\epsilon$, so $|w(\mathbf{x}) - P_o(\mathbf{x})| < \epsilon$. A similar argument can be applied to show the result (b) regarding the even function v. This proves the claim in Proposition 1. □

7.2 Proof of Proposition 2

Without loss of generality, we can set $\xi = 0$, $\sigma^2 = 1$, assume $\beta > 0$, and only need to prove that $\psi(x) = 2\phi(x)\Phi(\alpha x + \beta x^3)$ has at most two modes. We prove this by contradiction. If $\psi(x)$ has more than two modes, then $\psi'(x)$ has at least five zeros. In the following proof, we show that this cannot be the case.

We have $\psi'(x) = 2\phi(x)((\alpha + 3\beta x^2)\phi(\alpha x + \beta x^3) - x\Phi(\alpha x + \beta x^3))$ and need to consider 3 cases:

case 1: $\alpha = 0$

We write $\psi'(x) = 2x\phi(x)\eta(x)$, where $\eta(x) = 3\beta x\phi(\beta x^3) - \Phi(\beta x^3)$. We can verify that $\eta'(x) = 3\beta\phi(\beta x^3)\eta_1(y)$ where $y = x^2$ and $\eta_1(y) = 1 - y - 3\beta^2 y^3$. Since $\eta_1(y)$ is a decreasing function on $y \geq 0$, $\eta'(x)$ has at most two zeros. Thus, $\eta(x)$ has at most three zeros, hence $\psi'(x)$ has at most four zeros.

case 2: $\alpha > 0$

Notice that $\psi'(x) > 0$ for $x \leq 0$. For $\gamma_1(x) = \psi'(x)/(2x\phi(x)) = \phi(\alpha x + \beta x^3)(\alpha + 3\beta x^2)/x - \Phi(\alpha x + \beta x^3)$, we get $\gamma_1'(x) = \phi(\alpha x + \beta x^3)/(-9\beta x^2)\gamma_2(y)$, where $y = \alpha + 3\beta x^2 > 0$ and $\gamma_2(y) = y^4 + \alpha y^3 + (3 - 2\alpha^2)y^2 - (3\alpha + 9\beta)y + 18\alpha\beta$. Since $\gamma_2''(y) = 12y^2 + 6\alpha y + (6 - 4\alpha^2)$ has at most 1 positive zero, and $\gamma_2'(y) = 4y^3 + 3\alpha y^2 + (6 - 4\alpha^2)y - (3\alpha + 9\beta) < 0$ at $y = 0$, we know that $\gamma_2'(y)$ has at most one positive zero. Thus $\gamma_2(y)$ has at most 2 positive zeros. This means $\gamma_1'(x)$ has at most two positive zeros, so $\psi'(x)$ has at most three (positive) zeros.

case 3: $\alpha < 0$

Notice that $\psi'(x) < 0$ for $x \in [0, \sqrt{-\alpha/(3\beta)}\ ]$ and $\psi'(x) > 0$ for $x \in (-\infty, -\sqrt{-\alpha/(3\beta)}\ ]$. So we only look for solutions $x \in (\sqrt{-\alpha/(3\beta)}, \infty)$ and $x \in (-\sqrt{-\alpha/(3\beta)}, 0)$. Let $y = \alpha + 3\beta x^2$, then there is a one to one mapping between the x in the above range and $y \in (\alpha, \infty)$. Let $\gamma_1(x)$ and $\gamma_2(y)$ have the same expressions as in case 2. We have that $\gamma_2(y)$ has at most four zeros since it is a fourth order polynomial. Notice that $\gamma_2(\alpha) < 0, \gamma_2(-\infty) > 0$, so $\gamma_2(y)$ has at most three zeros in (α, ∞). This means $\gamma_1'(x)$ has at most three zeros, hence $\psi'(x)$ has at most four zeros. □

7.3 Proof of Proposition 3

Without loss of generality, we set $\boldsymbol{\xi} = \mathbf{0}$. The cdf of $\mathbf{x}$ is simply:

$$\begin{aligned} P(\mathbf{x} \le \mathbf{x}^*) &= P(\mathbf{y} \le \mathbf{x}^* | w(\mathbf{y}) > Z) \\ &= P(\mathbf{y} \le \mathbf{x}^*, w(\mathbf{y}) > Z)/P(w(\mathbf{y}) > Z) \\ &= 2\int_{-\infty}^{\mathbf{x}^*}\int_{-\infty}^{w(\mathbf{y})} f(\mathbf{y})dH(z)d\mathbf{y} \\ &= 2\int_{-\infty}^{\mathbf{x}^*} f(\mathbf{y})H(w(\mathbf{y}))d\mathbf{y}, \end{aligned} \tag{14}$$

and thus, by differentiating (14) with respect to $\mathbf{x}^*$, the result follows. Recall that w is odd, hence either $w(\mathbf{y}) > Z$ or $w(-\mathbf{y}) \ge -Z$, and thus $P(w(\mathbf{y}) > Z) = 1/2$. □

8 Reference

Arnold, B. C. and Beaver, R. J. (2000). The skew-Cauchy distribution, *Statistics and Probability Letters*, **49**, 285-290.

Arnold, B. C. and Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting, *Test*, **11**, 7-54.

Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scand. J. Stat.*, **12**, 171-178.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution, *Biometrika*, **83**, 715-726.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution, *J. R. Statist. Soc. B*, **61**, 579-602.

Azzalini, A. and Capitanio, A. (2002). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, *J. R. Statist. Soc. B*, accepted for publication.

Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions, *J. Multivariate Analysis*, **79**, 99-113.

Field, C., Genton, M. G. (2002). The multivariate g-and-h distribution, submitted to *Technometrics.*

Gallant, A. R. and Nychka, D. W. (1987). Seminonparametric maximum likelihood estimation, *Econometrica*, **55**, 363-390.

Genton, M. G., He, L. and Liu, X. (2001). Moments of skew-normal random vectors and their quadratic forms, *Statistics and Probability Letters*, **51**, 319-325.

Genton, M. G. and Loperfido, N. (2002). Generalized skew-elliptical distributions and their quadratic forms, *Institute of Statistics Mimeo Series #2539*, under review. (*http://www.stat.ncsu.edu/~mggenton/publications.html*)

Hoaglin, D. C. (1983). Summarizing Shape Numerically: The g-and-h Distributions. *Data Analysis for Tables, Trends and Shapes: Robust and Exploratory Techniques*, edited by D.C. Hoaglin, F. Mosteller, J.W. Tukey, Wiley, New York.

Jones, M. C., Faddy, M. J. (2002). A skew extension of the t distribution, with applications. Under revision.

Rudin, W. (1973). *Functional Analysis*, McGraw-Hill, Inc.

Sahu, S. K., Dey, D. K. and Branco, M. D. (2001). A new class of multivariate skew distributions with applications to Bayesian regression models, manuscript.

Wang, J. and Genton, M. G. (2002). The multivariate skew-slash distribution, manuscript.

Wang, J., Boyer, J. and Genton, M. G. (2002). A skew-symmetric representation of multivariate distributions, *Institute of Statistics Mimeo Series #2540*, under review. (*http://www.stat.ncsu.edu/~mggenton/publications.html*)